

Compte rendu - Modélisation par homologie d'un
domaine de la protéine p63 chez la souris.
Bio-informatique structurale

AMMICHE Naïma
M1 BI-IPFB, 2022-2023

Sommaire

Introduction :	2
Matériels et méthodes :	2
Résultats et discussion :	5
Identification des supports :	5
Alignement de la séquence cible avec les séquences supports :	7
Construction du modèle :	9
Evaluation du modèle :	10
Conclusion :	12
Bibliographie :	13

Introduction :

La protéine d'intérêt p63 est connue chez divers organismes dont la souris, le rat et l'homme. C'est une protéine qui agit comme un activateur ou un répresseur transcriptionnel de liaison à l'ADN spécifique à la séquence. Appartenant à la famille antigène de tumeur cellulaire p53, protéine tumorale 63 et appartient au gène TP63. Cette protéine possède divers noms comme Tumor protein 63, p63, Chronic ulcerative stomatitis protein (CUSP) ou encore Keratinocyte transcription factor KET. La structure de p63 est localisée au niveau du noyau. La structure 3D de p63 humaines est connue composer de 680 acides aminés, elle possède différents domaines et elle est composée d'hélices alpha et de boucles. Cependant la structure de cette protéine chez la souris et le rat n'est pas encore connue et uniquement une prédiction de la structure par alphafold est proposée. Cette protéine p63 peut causer des maladies comme le syndrome acro-dermo-ungual-lacrimal-tooth syndrome (ADULT syndrome) cette maladie cause l'ectrodactylie, une syndactylie, une dysplasie des ongles des doigts et des orteils, une atrésie du canal lacrymal, une alopecie frontale et la perte des dents permanentes. La protéine p63 peut également être responsable de Ankyloblepharon-défauts ectodermiques - fente labiale/palatine (AEC) responsable d'une hypoplasie maxillaire, une hypodontie et une fente labio-palatine. La souris étant un organisme étudié en laboratoire, la structure 3D de la protéine p63 chez la souris peut être utilisée pour la recherche de médicaments potentiels pour les maladies causées par le p63. L'objectif de ces TP est de réaliser une modélisation par homologie afin de modéliser un domaine de la p63 chez la souris en utilisant successivement un support p63, un support p73 et un support p53 à l'aide d'outils tels que uniprot, blast, pymol et MODELLER software. La démarche à suivre consiste à identifier les supports qui serviront de squelette pour la prédiction, réaliser un alignement de la séquence cible et la séquence support pour construire le fichier au format ALI qui permet ensuite de construire le modèle, enfin, une évaluation du modèle est réalisée. Le fichier au format ALI contient l'alignement de séquences entre la séquence cible et la séquence support.

Matériels et méthodes :

Nous disposons de la séquence cible p63 chez la souris qui agit comme un activateur ou répresseur transcriptionnel de liaison à l'ADN, elle peut initier l'apoptose dépendante de

p53/TP53, active la transcription du promoteur p21 et active la transcription de RIPK4. Elle joue un rôle dans la régulation de la morphogenèse épithéliale. Nous réalisons une modélisation par homologie. La modélisation par homologie est divisée en quatre étapes pour construire et évaluer un modèle fiable pour un domaine de la protéine p63 chez la souris. Premièrement, l'identification des supports qui serviront de squelette pour la prédiction, ces supports sont sélectionnés via différents critères comme la similarité de séquence avec la séquence cible de la protéine p63 de souris, la disponibilité et qualité des structures cristallographiques ou RMN pour les différentes séquences de support. La base de données de séquences de protéine uniprot est utilisée, elle permet de recueillir des informations concernant les fonctions, la structure d'une protéine d'intérêt. Les séquences peuvent être review ou non review c'est-à-dire annoté ou non manuellement. Nous ouvrons uniprot et tapons dans la barre de recherche p63, sélectionner review (à gauche) et ensuite l'identifiant O88898 correspondant à la protéine p63 chez la souris, les diverses sections de cette page permettent de recueillir diverses informations sur cette protéine. La section « Sequences & isoforms » puis download permet de récupérer la séquence protéique. Il s'agit de la séquence cible, nous entrons ensuite cette séquence BLAST¹ dans la section saisir la séquence FASTA, BAST permet de réaliser des alignements locaux pour trouver des régions similaires entre les séquences biologiques. Les paramètres à entrer sont standard databases, Protein Data Bank proteins(pdb) pour choisir un support dont la structure tridimensionnelle est déterminée expérimentalement. PSI-BLAST qui est un algorithme de recherche de similarités de séquences dans des bases de données protéiques qui permet de détecter les séquences de protéines éloignées dans les bases de données puis appuyées sur BLAST.

The screenshot shows the NCBI BLAST search interface. The 'Choose Search Set' section has 'Standard databases (nr etc.)' selected. The 'Database' dropdown is set to 'Protein Data Bank proteins(pdb)'. The 'Organism' field is empty. The 'Exclude' section has 'Models (XM/XP)' checked. The 'Program Selection' section shows 'PSI-BLAST (Position-Specific Iterated BLAST)' selected.

Figure 1: Paramètre à entrer pour réaliser un BLAST sur la séquence de la p63 chez la souris.

Une page chargera et contiendra l'ensemble des résultats, un second blast est lancé pour cela, il faut décocher la sélection de la première colonne (à gauche) en décochant select all et cocher uniquement la colonne select for PSI blast (en bas à droite) puis RUN. Les résultats obtenus permettent ensuite de sélectionner les supports adapter selon leurs pourcentages d'identité avec la séquence cible. Lorsque les supports sont sélectionnés, leur fichier au format PDB est récupéré en utilisant RCSB PDB, en entrant dans la barre de recherche le

¹ [Protein BLAST: search protein databases using a protein query \(nih.gov\)](https://blast.ncbi.nlm.nih.gov/Blast.cgi)

code associé au support choisi. Les séquences des supports au format FASTA sont récupérées à partir des fichiers au format PDB à partir d'un site de conversion de format².

La seconde étape consiste à réaliser un alignement de la séquence cible avec la séquence des supports sélectionnés pour obtenir le fichier au format ALI, via EMBOSS Needle qui est un programme d'alignement global de séquences utilisé pour comparer deux séquences d'acides aminés ou de nucléotides. Il utilise l'algorithme de Needleman-Wunsch pour aligner les séquences. Les paramètres nécessaires sont deux séquences à aligner donc la séquence cible et la séquence support, en types de séquence mettre PROTEIN, en output format mettre pair, en matrix mettre blosum62 qui est pertinent pour l'alignement de séquences protéiques, pénalité d'ouverture de gap à 10 et pénalité d'extension de gap à 0.5 pour maximiser la qualité de l'alignement. Réaliser un second alignement en modifiant uniquement l'output format à Pearson/FASTA afin d'obtenir un alignement global où les séquences possèdent la même taille. Les paramètres sont donc les suivants.

STEP 2 - Set your pairwise alignment options

OUTPUT FORMAT

pair

MATRIX	GAP OPEN	GAP EXTEND	END GAP PENALTY	END GAP OPEN	END GAP EXTEND
BLOSUM62	10	0.5	false	10	0.5

Figure 2 : Paramètres à entrer sur EMBOSS Needle pour réaliser un alignement global à partir de la séquence p63 chez la souris et les séquences des supports.

L'alignement local est ensuite réalisé en utilisant EMBOSS Water qui est également un programme d'alignement de séquences similaire à EMBOSS Needle qui permet de réaliser un alignement local, mais qui utilise l'algorithme Smith-Waterman pour aligner les séquences. Les paramètres à entrer sont similaires à ceux de EMBOSS Needle.

STEP 2 - Set your pairwise alignment options

OUTPUT FORMAT

pair

MATRIX	GAP OPEN	GAP EXTEND
BLOSUM62	10	0.5

Figure 3 : Paramètres à entrer sur EMBOSS Water pour réaliser un alignement local à partir de la séquence p63 chez la souris et les séquences des supports.

Ensuite nous avons procédé à un alignement par paires optimales pour améliorer la qualité de l'alignement initial obtenu par EMBOSS Needle ou Water et confronté cet alignement à un alignement multiple (MSA) pour nous assurer de la qualité de l'alignement. L'alignement par paires optimales est réalisé avec Clustal Omega qui est un logiciel d'alignement de séquences multiple, il utilise l'algorithme de guide-arbre pour optimiser l'alignement. Les paramètres nécessaires sont plusieurs séquences à aligner, en types de séquence PROTEIN et les paramètres suivants.

² [pdb2fasta \(zhanggroup.org\)](http://pdb2fasta.zhanggroup.org)

STEP 2 - Set your parameters

OUTPUT FORMAT
Pearson/FASTA

DEALIGN INPUT SEQUENCES	MBED-LIKE CLUSTERING GUIDE-TREE	MBED-LIKE CLUSTERING ITERATION	NUMBER of COMBINED ITERATIONS
no	yes	yes	default(0)
MAX GUIDE TREE ITERATIONS	MAX HMM ITERATIONS	DISTANCE MATRIX	GUIDE TREE
default	default	no	yes

ORDER
aligned

Figure 4 : Paramètres à entrer sur Clustal Omega pour réaliser un alignement multiple à partir de la séquence p63 chez la souris et les séquences des supports.

La troisième étape nécessite l'utilisation du logiciel de modélisation moléculaire modeller qui permet la construction de modèles protéiques à partir de leurs séquences d'acides aminés. Des techniques de modélisation par homologie afin de prédire la structure d'une protéine inconnue en utilisant une structure connue d'une protéine similaire. Modeller est donc utilisé pour construire les modèles à partir des fichiers au format ALI contenant la séquence cible et la séquence des supports, le format PDB des supports et le script python pour chaque support sélectionné. La qualité de chaque modèle est ensuite vérifiée à l'aide des scores de qualité molpdf et DOPE permettant de déterminer le modèle le plus fiable. Le script python contient les instructions, paramètres et fichiers permettant de générer des modèles. Les fichiers au format PDB contiennent les structures 3D des protéines supports.

Enfin, l'évaluation de chaque modèle avec verify 3D et ProSA-Web est réalisé. Nous avons également comparé la qualité des différents supports utilisés, en utilisant les résultats du CASP14. L'évaluation finale de chaque modèle est déterminé avec ProSA-Web. Ce protocole permet la construction et évaluation de modèle pour un domaine de P63 de souris en utilisant divers supports.

Résultats et discussion :

Identification des supports :

Les résultats de PSI-BLAST de la protéine p63 chez la souris sont les suivants.

Descriptions	Graphic Summary	Alignments	Taxonomy	Download	Select columns	Show	500
Sequences producing significant alignments							
110 sequences selected							
Sequences with E-value BETTER than threshold							
select all 110 sequences selected Skip to the first new sequence							
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per ident	Acc. Len
Crysto-EM structure of p200-p53 protein complex [Homo sapiens]	Homo sapiens	466	466	38%	2e-161	49.62%	265
Chain K: Cellular tumor antigen p53 [Homo sapiens]	Homo sapiens	455	455	38%	3e-155	49.62%	395
Chain M: Cellular tumor antigen p53 [Homo sapiens]	Homo sapiens	455	455	38%	3e-155	49.62%	393
HUMAN p53 CORE DOMAIN MUTANT M133L-V203A-Y239F-N239Y-T253I-N268D [Homo sapiens]	Homo sapiens	415	415	32%	2e-142	54.30%	219
Crystal structure of the human p53 core domain mutant M133L-V203A/N239Y/N268D at 1.9 Å resolution [Homo sapiens]	Homo sapiens	413	413	32%	1e-141	53.39%	219
p53 CORE DOMAIN IN COMPLEX WITH DNA [Homo sapiens]	Homo sapiens	412	412	32%	2e-141	52.94%	219
The solution structure of the p53 DNA-binding domain [Homo sapiens]	Homo sapiens	413	413	34%	2e-141	97.46%	233
Human p53 Core Domain Mutant N239K [Homo sapiens]	Homo sapiens	412	412	32%	4e-141	52.49%	219
Crystal structure of the BRCT domains of 53BP1 in complex with p53 and H2AX-pSer139 [Homo sapiens]	Homo sapiens	412	412	32%	5e-141	52.73%	225
Human p53 Core Domain Mutant V157E [Homo sapiens]	Homo sapiens	412	412	32%	5e-141	52.94%	220
An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity [Homo sapiens]	Homo sapiens	412	412	38%	5e-141	46.59%	234
Solution Structure of the Complex Between BCL-xL and the p53 Core Domain determined with P [Homo sapiens]	Homo sapiens	411	411	32%	5e-141	52.97%	217
Human p53 Core Domain Mutant N239Y [Homo sapiens]	Homo sapiens	411	411	32%	6e-141	52.49%	219
An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity [Homo sapiens]	Homo sapiens	412	412	38%	6e-141	46.56%	231
p53 DNA binding domain in complex with Np139 [Homo sapiens]	Homo sapiens	411	411	32%	8e-141	52.94%	227

Figure 5 : Représentation des résultats de blast sur la séquence de la p63 chez la souris. Avec les paramètres décrit précédemment.

Deux itérations de PSI-BLAST sont réalisées, nous observons différentes protéines avec une description, l'organisme, le score maximal, le score total, le query cover qui indique la proportion de la séquence de la protéine d'intérêt qui est couverte par l'alignement avec la séquence de référence. Lorsque la couverture est élevée cela signifie que la similarité entre les séquences est importante. Nous avons E-value lorsqu'il est significatif alors l'homologie de séquence n'est pas dû au hasard, le pourcentage d'identité, la taille de la séquence protéique et le code d'accension. Le choix des supports est important pour réaliser la modélisation par homologie pour cela nous sélectionnons des supports dont le pourcentage d'identité est supérieur à 40% et qui n'est pas de 100%, avec un E-value significatif. Cliquer sur l'accension ou en entrant le code d'accension du support dans RCSB PDB, nous avons des informations concernant le support choisi. Nous sélectionnons les supports suivants.

Pourcentage de similarité avec la séquence d'intérêt	Code d'accension du support choisi	Longueur de la séquence protéique (en acide aminé)	Chaîne	Nom de la protéine	Méthode	Résolution
97,46%	2RMN_A	233	A	Tumor protein 63	RMN	
52,94%	4QO1_B	227	B	Cellular tumor antigen p53	Rayon X	1.92 Å
81,16%	2XWC_A	208	A	TUMOUR PROTEIN P73	Rayon X	1,82 Å

Les supports choisis ont un pourcentage de similarité avec la séquence cible supérieure à 40%. Leur structure tridimensionnelles est obtenue par rayon X pour la chaîne B du domaine de liaison à l'ADN de p53 en complexe avec Nb139 avec une résolution de 1,92 angströms (4QO1_B) et la chaîne A de la structure cristalline du domaine de liaison à l'ADN du TP73 humains avec une résolution de 1,82 angströms (2XWC). La structure en solution du domaine de liaison à l'ADN de la p63 (2RMN_A) est obtenue par RMN, la p63-DBD peut se lier à l'anti-apoptotique BclxL via son interface de liaison à l'ADN.

Nous récupérons ensuite les formats PDB des supports sur RCSB PDB en entrant le code d'accèsion puis download Files puis PDB format. Le site de conversion est ensuite utilisé comme indiqué précédemment pour récupérer à partir du fichier PDB la séquence protéique au format FASTA associé.

Alignement de la séquence cible avec les séquences supports :

Un alignement global avec EMBOOS Needle est réalisés sur la séquence d'intérêt de la protéine p63 chez la souris et les supports sélectionnés. Cet alignement permet d'aligner la séquence complète de la protéine d'intérêt avec celle des protéines support pour déterminer précisément les positions des résidus et permettre la construction d'une structure tridimensionnelle de la protéine support comme référence.

```
>P63_MOUSE O88898 Tumor protein 63 OS=Mus musculus OX=10090 GN=TP63 PE=1 SV=1
MHFETSRCATLQYCPDPIYQIFETPAHFSAKESYRSMSQSTQSEFLSPEVFQHTWD
FLEQPCISVQPIELNFVDEPSENGATNKEIISMDICIRMQSDLSDPHMPQVTHLGLNSM
DQQTQNGSSSTSPYNTDHAQNSVTAPSPYAQPSSTFDALSPSPAIPSNIDVPGHSDVS
FQQSSAKSATMTYSTELKKLYCQIAKTCPIQIKVMTPPPQAVTRAMPVYKKAHVTEV
VKRCPIHELSREFNEGQIAPPSSLIRVEGNHQAQVEDPITGRQSVLPVPEPPQVTEFT
TVLYNFMCHNSCVGGMRRPILIIIVTLETRDQVLRRCFEARICACPGDRKADEDSIR
KQQVSDSAKNGDGTGRFRQNTHGIMTISIKRRSPDELLYLPVRGRETYEMLKIKES
LELNQYLPQHTIETYYQQQQQQHLLQKQTSNQSSSYGNSPPLHKNHSHNKLPSVSQ
LTNPQRNALTPITNPEGNRIIPMTNHPAGDMGLSPQALPPPLSPPSTSHCTPP
PPVPTDCSIVSFLARLGCSSCLDYFTTQGLTTIVQIEHYSHDGLASKIPEQFRHAIKKG
ILDHRQLHDFSSPPHLLRTPSGASTVSGSSETRGERVIDAVRFTLRQTISFPDRDENND
FNFDHDSRRNKQQRKEEGE
>2RMN_A
-----
-----GSSTFDALSPSPAIPSNIDVPGHSDVS
FQQSSAKSATMTYSTELKKLYCQIAKTCPIQIKVMTPPPQAVTRAMPVYKKAHVTEV
VKRCPIHELSREFNEGQIAPPSSLIRVEGNHQAQVEDPITGRQSVLPVPEPPQVTEFT
TVLYNFMCHNSCVGGMRRPILIIIVTLETRDQVLRRCFEARICACPGDRKADEDSIR
KQQVSDSTKNGDA-----FRQNTHGIMT-----
-----
-----
```

Figure 6 Alignement global entre la séquence cible de la souris et la séquence de la chaîne A de la protéine tumorale 2RMN

Nous observons que les gaps sont majoritairement placés aux extrémités de la séquence support. Le domaine DNA binding est présent au milieu de la séquence cible. Un profil similaire est obtenu pour l'alignement global de la protéine d'intérêt p63 chez la souris avec respectivement les protéines supports 4QO1_B correspondant à la chaîne B de l'anti-gène tumoral cellulaire p53 et 2XWC_A correspondant à la chaîne A de la protéine tumoral p73.

L'alignement local avec emboss water permet d'avoir des informations concernant le pourcentage d'identité, de similarité, de gap et le score. Cet alignement local permet d'identifier les régions de la séquence d'intérêt qui ont une similarité de séquence significative avec les séquences supports. Ces régions correspondent aux régions conservées ou régions homologues. L'identification de ces régions permet de construire un modèle 3D plus précis. Cet alignement permet également d'identifier les régions qui ne sont pas homologues avec la séquence support, et identifier ainsi les zones du modèle qui sont incertaines lors de

Q001_B	sp 088898 P63_MOUSE	0
2R9M_A	0	
2XK_A	0	
Q001_B	sp 088898 P63_MOUSE	0
2R9M_A	120	
2XK_A	0	
Q001_B	sp 088898 P63_MOUSE-PLSSVSQ:TVQSGVSGILG	180
2R9M_A	180	
2XK_A	29	
Q001_B	sp 088898 P63_MOUSEAPVTSINTDYGPHFIVT	190
2R9M_A	190	
2XK_A	0	
Q001_B	sp 088898 P63_MOUSE	FLHSGTAISVTTGPSALNHQFQAICTPQGLVDSPTPGTIVIAHATVYQSGHTEV	81
2R9M_A	240	
2XK_A	79	
Q001_B	sp 088898 P63_MOUSE	VHPCPHHLCSD--SDGLAPQHLTVEGNLIVELYDQMTIHSHVVPVPPVGGSDCT	130
2R9M_A	130	
2XK_A	139	
Q001_B	sp 088898 P63_MOUSE	TLHYVKNSSCGGWNRPILITLLEDSQILLGNSEFVIVCACPGDRIITEENL	199
2R9M_A	199	
2XK_A	209	
Q001_B	sp 088898 P63_MOUSE	TLHYVKNSSCGGWNRPILITLLETHQDQGLCEFAITCACPGDRIITEENL	199
2R9M_A	199	
2XK_A	209	
Q001_B	sp 088898 P63_MOUSE	IQQVSDSANGDGTTPPTQITLKEIGTISIKRRSPDELLVLPVYGTIEHLLIKES	233
2R9M_A	233	
2XK_A	284	
Q001_B	sp 088898 P63_MOUSE	LELKLVLPHQITELYVQQQQQKHLQLQTSQSGQSVKNSSPPLINYSNPLPSVQ	199
2R9M_A	233	
2XK_A	284	
Q001_B	sp 088898 P63_MOUSE	LTNPQKIALTPPTLPGNGANPTPGTHTPMAGDNLSPQALPPLSPMSTSHCTPT	199
2R9M_A	240	
2XK_A	284	
Q001_B	sp 088898 P63_MOUSE	PPPTDCISVSFLAILGCSSCLDYFTQQLTTIVQITLHMSNDOLASLKEPQRIHAIHG	0
2R9M_A	0	
2XK_A	233	
Q001_B	sp 088898 P63_MOUSE	ILIHQLHRSFSPHLLITLPGSTVSGGSETEGEIVDAVITLTQTISPPPIIDEN	666
2R9M_A	233	
2XK_A	284	
Q001_B	sp 088898 P63_MOUSE	199
2R9M_A	680	
2XK_A	284	

Nous observons que les séquences supports ont majoritairement des gaps aux extrémités de leur séquence protéiques ce qui témoigne que le domaine DNA binding n'est pas présent aux extrémités de la séquence cible. Une région conservée correspondant au domaine DNA binding est visible entre les résidus 170 à 362 de la séquence cible.

Nous observons que les séquences supports ont majoritairement des gaps aux extrémités de leur séquence protéiques ce qui témoigne que le domaine DNA binding n'est pas présent aux extrémités de la séquence cible. Une région conservée correspondant au domaine DNA binding est visible entre les résidus 170 à 362 de la séquence cible.


```

1  #!/usr/bin/env python
2  # Homology modeling by the automodel class
3  from modeller import *          # Load standard Modeller classes
4  from modeller.automodel import * # Load the automodel class
5  log.verbose()                  # request verbose output
6  env = environ()                # create a new MODELLER environment to build this model in
7  # directories for input atom files
8  #env.io.atom_files_directory = './../atom_files'
9  #env.io.atom_files_directory = '/home/../../modeller'
10
11 #Read in HETATM records from template PDBs
12 env.io.hetatm = True
13
14 a = automodel(env,
15               alnfile = 'alignementDHH.ali', # alignment filename
16               knowns = ('2rmn','2xwc','4qo1'), # codes of the templates
17               sequence = 'p63', # code of the target
18               assess_methods=(assess.DOPE, assess.GA341))
19 a.starting_model= 1 # index of the first model
20 a.ending_model =100 # index of the last model
21 # (determines how many models to calculate)
22 a.make() # do the actual homology modeling

```

Figure 10: script python à utiliser.

Nous lançons Modeller avec la commande **mod10.4 script.py** afin de générer 100 modèles.

Evaluation du modèle :

Parmi les 100 modèle générer, nous sélectionnons le meilleur. Nous récupérons dans une fichier nommer bs_score les informations des 100 modèles générer tel que.

	Filename	molpdf	DOPE score	GA341 score
1	p63.B99990001.pdb	11884.31152	-32610.13281	1.00000
2	p63.B99990002.pdb	11518.14258	-33082.62109	1.00000
3	p63.B99990003.pdb	11452.18457	-32540.25000	1.00000
4	p63.B99990004.pdb	11689.01855	-32766.95508	1.00000
5	p63.B99990005.pdb	11872.75879	-32618.39453	1.00000
6	p63.B99990006.pdb	11855.72363	-32707.75977	1.00000
7	p63.B99990007.pdb	11404.69238	-32945.19922	1.00000
8	p63.B99990008.pdb	11621.36035	-32677.03125	1.00000
9	p63.B99990009.pdb	11617.82520	-33008.64453	1.00000

Figure 11: Visualisation des informations des 9 premiers modèles générer par Modeller à partir du script python décrit précédemment. Résultat récupérer dans le fichier bs_score avec en ligne les modèles, et en colonne le nom du fichier pdb, le molpdf, le score DOPE ainsi que le score GA341 de chaque modèle générer par Modeller.

Nous sélectionnons le score DOPE minimum, ce qui correspond à l'énergie la plus basse du modèle. Le script R suivant permet de récupérer le numéro du modèle qui possède un score DOPE minimum.

```

data<-read.delim("bs_score", sep = "")
min_score = min(data$DOPE)
data[which(data$DOPE == min_score), ]

```

Nous évaluons ensuite ce modèle qui possède le score DOPE minimum avec les logiciels PROSA-WEB et VERIF3D.

Z-Score: -3.57

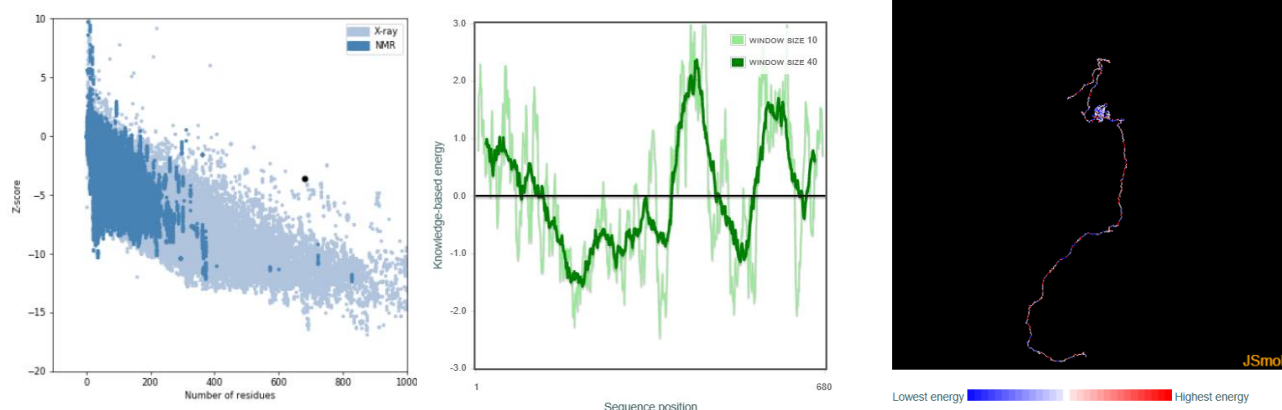


Figure 12: Evaluation du 38eme modèle généré par Modeller à l'aide de PROSA-WEB. Le graphique 1 représente le Z-score en fonction du nombre de résidus avec en bleu clair le rayon X et en bleu foncé le RMN. Le seconde graphique représente l'énergie en fonction de la position des résidus. Le troisième représente la structure tridimensionnel du modèle généré avec en bleu les zone de faible énergie et en rouge de forte énergie.

Un Z-score idéal est d'environ 0, ce qui indique que le modèle a une qualité similaire à celle d'une structure expérimentale résolue à haute résolution. Les z-scores peuvent varier selon de résolution de la structure, la taille de la protéine. Nous observons un Z-score de -3,57 ce qui indique que le modèle peut avoir des problèmes de qualités. Concernant le second graphique, les valeurs positives correspondent à des parties problématiques du modèle. Le modèle possède d'après le second graphique une énergie haute pour les résidus 1 à 170 et de 362 à 680 il s'agit des zones mal repliées. L'énergie du modèle est faible pour les résidus au position 170 à 362. Notre modèle possède une structure tridimensionnelle de bonne qualité pour cette région de faible énergie. La représentation du modèle montre que les régions de haute énergie correspondent au boucle, les régions avec de faible énergie correspondent au hélice alpha et au brin bêta.

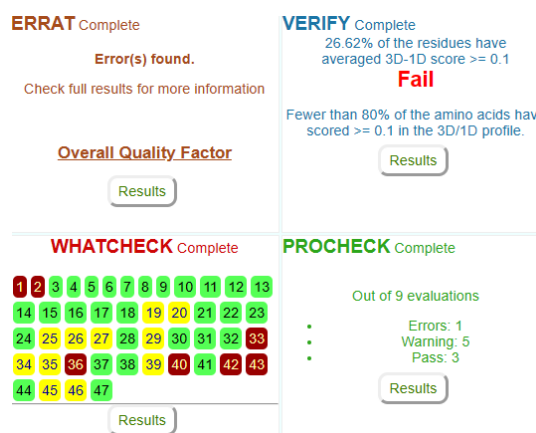


Figure 13 : Résultat de l'évaluation du modèle avec Verify 3D avec les résultats de Verify, whatcheck et procheck.

Verify 3D permet d'analyser la compatibilité du modèle atomique avec sa séquence en acides aminés. Nous avons 26,62% des résidus qui ont un score 3D-1D moyen supérieur ou égal à 0.1 ce qui est faibles. Car moins de 80% des acides aminés ont obtenu un score supérieur ou égal à 0.1 dans le profil 3D/1D. Ce modèle permet d'avoir 26,62% de la

structure tridimensionnel de la séquence cible. La région du DNA binding est de taille 192 résidus (du résidu 170 à 362) et la protéine p63 de souris possède 685 résidus, ainsi la région DNA binding correspond à 28% de la structure tridimensionnel. Les résultats de ces programmes indiquent que le modèle de la protéine cible n'est pas très fiable et qu'il contient des erreurs et des avertissements. Cependant, ces programmes fournissent également des informations utiles sur certaines parties de la séquence de la protéine cible, telles que les régions qui ont une conformation 3D incorrecte. Cela peut être utile pour les chercheurs qui souhaitent améliorer le modèle en corrigeant ces régions problématiques.

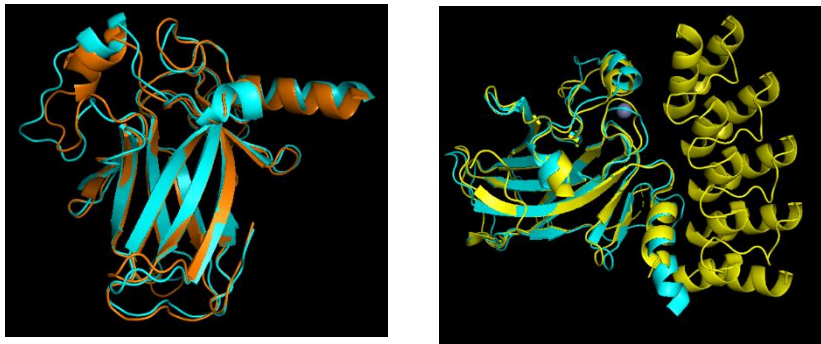


Figure 14 : Représentation du modèle du domaine DNA binding de la p63 chez la souris en cyan, le domaine de la DNA binding (170 à 362) de la p63 de souris du modèle alpha fold en orange, et en jaune la chaîne A du domaine DNA binding de la p63 résolu chez l'homme (code d'accèsion 7z7e). Un RMSD pour le modèle en cyan avec le modèle orange, puis avec le modèle jaune.

Le modèle du domaine DNA binding de la p63 chez la souris avec celui d'alphafold a permis d'obtenir un RMSD de 0,851. La chaîne A du domaine DNA binding de la p63 chez l'homme est résolu par rayon X avec une résolution de 1,80 angströms. Cette structure en cyan avec le domaine DNA binding de la p63 chez l'homme a permis d'obtenir un RMSD de 0,971. Les RMSD obtenus sont faibles ce qui signifie que la structure est proche. La structure du domaine de DNA binding de la p63 chez la souris est similaire à la chaîne A du domaine DNA binding chez l'homme. Une fonction similaire de ces deux domaines est donc possible.

Conclusion :

Le modèle du domaine DNA binding de la protéine p63 chez la souris permet de confirmer que celui-ci est localiser entres les résidus 170 et 362. La structure tridimensionnelle du modèle montre que cette région est composé de feuillet bêta anti parallèle, hélices alpha et de boucle. Cette structure est très similaire à la chaîne A du même domaine chez l'homme ce qui pourrait être une piste pour comprendre le rôle et fonctionnement de ce domaine. Ce projet a permis d'obtenir des résultats intéressant et utiles pour les futures études sur ce domaine de la protéine. Les supports montre que la structure en solution du p63-DBD est moins dynamique que celle de la p53. L'utilisation de modèle de prédiction pour la structure de la protéine p63 chez la souris ne permet pas une prédiction précise à 100%. Les modèles obtenus fournissent des information utiles pour guider la recherche expérimentale et permettent de mieux comprendre la structure et la fonction de la protéine p63 chez la souris. C'est pourquoi l'évaluation de la qualité des modèles obtenus est importante en utilisant notamment PROSA-WEB, PROCHECK et VERIF3D qui permettent d'avoir une idée de

fiabilité des résultats obtenus. Cet outils ne remplacent pas les techniques expérimentales pour la résolution de la structure de protéine mais permet de les compléter.

Bibliographie :

[TP63 - Tumor protein 63 - Homo sapiens \(Human\) | UniProtKB | UniProt](#)

[Tp63 - Tumor protein 63 - Mus musculus \(Mouse\) | UniProtKB | UniProt](#)

[Tp63 - Tumor protein 63 - Rattus norvegicus \(Rat\) | UniProtKB | UniProt](#)

[Protein BLAST: search protein databases using a protein query \(nih.gov\)](#)

[About MODELLER \(salilab.org\)](#)

[EMBOSS Water < Pairwise Sequence Alignment < EMBL-EBI](#)

[EMBOSS Needle < Pairwise Sequence Alignment < EMBL-EBI](#)